Adding an Annotation Model
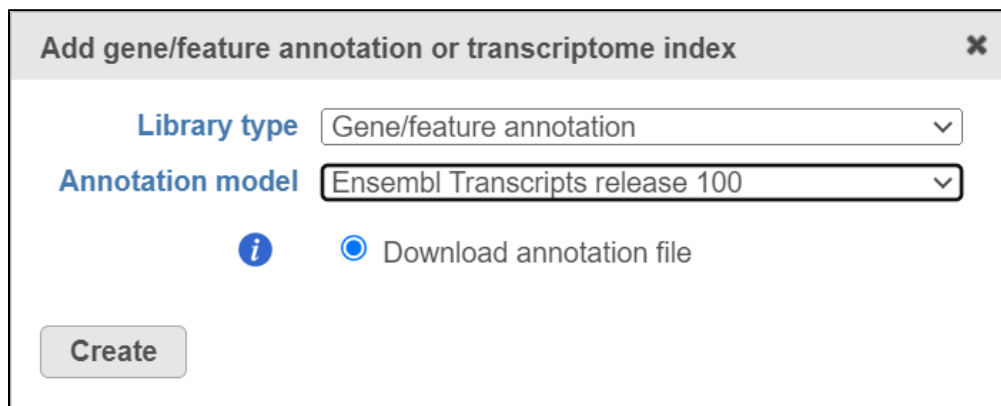
On the library file management page, click the **Add annotation models** green button in the *Annotation models* section, choose **Gene/feature annotation** from the *Library type* drop-down list in the dialog (Figure 1). Alternatively, click the **Add library file** button and choose **Gene/feature annotation** from the *Library type* drop-down list.
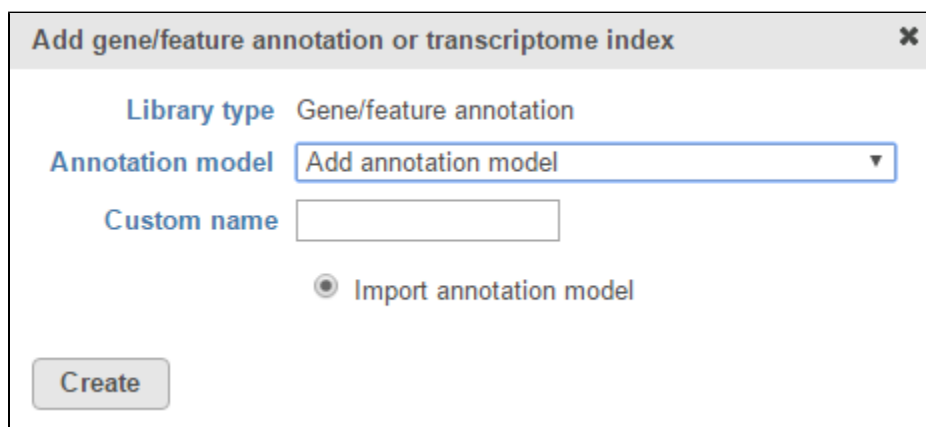
If you are using an assembly supported by Partek (e.g. human), annotation models from a variety of commonly used sources (e.g. ENSEMBL, GENCODE) will appear in the *Annotation model* drop-down list in the dialog. Choose an annotation model, select the **Download annotation file** radio button and click **Create** (Figure 1).

Figure 1. For many model organisms, automatic downloads of gene/feature annotation models from various sources are available from the Partek repository.

If you prefer to add a custom annotation file or if you are working with a custom assembly, choose **Add annotation model** from the *Annotation model* drop-down list (Figure 2). You may need to scroll to the bottom of the drop-down list to see this option. Name the annotation model by typing into the *Custom Name* box and click **Create**. Characters such as $ * | \ : " < > ? / % cannot be used in custom names.

Figure 2. Add a custom annotation model

A custom annotation model can be added from the *Partek Flow Server*, *My Computer* or a *URL* download link. The behavior of each option is similar to when importing a reference sequence (see *Adding a Reference Sequence*). When browsing for files on the Partek Flow server, only the files with relevant file extensions will be visible (.gtf, .gff, .gff3, .bed, .pannot, .txt and various compressed formats).

Select the source of annotation file and click *Next*. If the annotation file format is gtf, gff, gff3 or bed, a preview of the first 20 rows of the file will be shown on the screen (Figure 3). You must then specify the type of annotation file by choosing an option from the *Annotation type* drop-down list (Figure 3). Choose *mRNA* for whole transcriptome annotations, where both gene and transcript level information are present. Choose *microRNA* for precursor or mature microRNA transcript annotations. Choose *Amplicon* for targeted amplicon sequencing annotations. Choose *Other* for annotation files that do not fall into the other three categories (e.g. lncRNA).

Annotation type ⓘ   [mRNA ▾]

	-- Select --		
Column	**mRNA** ▸ ontent	**Transcript ID**	**Gene ID**
gene_id	microRNA 160072, ENSG0...	○	○
gene_version	Protein 0, 20, 20...	○	○
transcript_id	Amplicon ENST00000673477, ENST0...	○	○
transcript_version	Other 1, 1, 1, 1, 1, 1, 1, 1, 1	○	○
exon_number	1, 1, 2, 2, 3, 3, 4, 4, 5	○	○
gene_name	ATAD3B, ATAD3B, ATAD3B...	○	○
gene_source	ensembl_havana, ensemb...	○	○
gene_biotype	protein_coding, protei...	○	○
transcript_name	ATAD3B-206, ATAD3B-206...	○	○
transcript_source	ensembl_havana, ensemb...	○	○
transcript_biotype	protein_coding, protei...	○	○
tag	CCDS, basic, Ensembl_c...	○	○
ccds_id	CCDS30, CCDS30, CCDS30...	○	○
exon_id	ENSE00003889014, ENSE0...	○	○
exon_version	1, 1, 1, 1, 1, 1, 1, 1, 1	○	○
protein_id	ENSP00000500094, ENSP0...	○	○
protein_version	1, 1, 1, 1, 1, 1, 1, 1, 1	○	○
transcript_support_level	1, 1, 1, 1, 1, 1, 1, 1, 1	○	○

Show/hide file preview

[Back]   [Finish]

Figure 3. When importing an annotation model, the data type must be specified

For text annotations with both gene- and transcript-level information select *mRNA* from the drop-down menu. You can manually specify which column corresponds to the **Transcript ID** or **Gene ID** by selecting the corresponding radio button next to the column with that information (Figure 4).

Column	Content	Transcript ID	Gene ID
gene_id	ENSG00000160072, ENSG0...	○	◉
gene_version	20, 20, 20, 20, 20, 20...	○	○
transcript_id	ENST00000673477, ENST0...	◉	○
transcript_version	1, 1, 1, 1, 1, 1, 1, 1, 1	○	○
exon_number	1, 1, 2, 2, 3, 3, 4, 4, 5	○	○
gene_name	ATAD3B, ATAD3B, ATAD3B...	○	○
gene_source	ensembl_havana, ensemb...	○	○
gene_biotype	protein_coding, protei...	○	○
transcript_name	ATAD3B-206, ATAD3B-206...	○	○
transcript_source	ensembl_havana, ensemb...	○	○
transcript_biotype	protein_coding, protei...	○	○
tag	CCDS, basic, Ensembl_c...	○	○
ccds_id	CCDS30, CCDS30, CCDS30...	○	○
exon_id	ENSE00003889014, ENSE0...	○	○
exon_version	1, 1, 1, 1, 1, 1, 1, 1, 1	○	○
protein_id	ENSP00000500094, ENSP0...	○	○
protein_version	1, 1, 1, 1, 1, 1, 1, 1, 1	○	○
transcript_support_level	1, 1, 1, 1, 1, 1, 1, 1, 1	○	○

Figure 4. Select the column with gene- and transcript level information

Genome coordinates for annotation models stored in Partek Flow are 1-based, start-inclusive, and stop-exclusive. This means that the first base position starts from one, the start coordinate for a feature is included in the feature and the stop/end coordinate is not included in the feature. These are the genome coordinates that are printed in various task reports and output files when an annotation model is involved in the task. When custom annotation files are added to Partek Flow, the genome coordinates are converted into this format. The coordinates are converted back if necessary for a specific task. Figure 5 shows how the genome coordinates vary between different annotation formats.

Format	1- or 0-based	Start	Stop/End	Example (Start, Stop/End)
GTF/GFF/GFF3	1	inclusive	inclusive	1000, 2000
BED	0	inclusive	exclusive	999, 2000
Partek	1	inclusive	exclusive	1000, 2001

Figure 5. Comparison of genome coordinate formats. The example column shows the coordinates for a feature that starts at position 1000 (inclusive) and stops at position 2000 (inclusive).

Additional Assistance

If you need additional assistance, please visit our support page to submit a help ticket or find phone numbers for regional support.

« Adding a Variant Effect Predictor (VEP) Database   Adding Aligner Indexes Based on an Annotation Model »

Your Rating: ☆☆☆☆☆   Results: ★★★★⯪ 41 rates